Assessment of Mid-Range Computing at LBNL

Study sponsored by CSAC and ITSD

Paul D. Adams, Physical Biosciences
Jon Bashor, Computing Sciences
Ali Belkacem, Chemical Sciences
Alessandra Ciocio, Physics
Kenneth H. Downing, Life Sciences
Gary Jung, Information Technologies and Services
James F. Leighton, Information Technologies and Services
Alexander "Sandy" Merola, Information Technologies and Services
Douglas L. Olson, Nuclear Science
John W. Staples, Accelerator and Fusion Research
Shaheen Tonse, Environmental Energy Technologies
Michel A. Van Hove, Materials Sciences
Tammy S. Welcome, NERSC

Abstract

As the role and contributions of high-performance computing continue to increase in significance, Berkeley Lab scientists have been seeking out potential advantages provided by more powerful computing resources. This report documents the current status of mid-range computing at Berkeley Lab and describes the process set out by the Computing and Communication Services Advisory Committee (CSAC) and the Information Technologies and Services Division (ITSD) to study ways of enhancing mid-range computing at the Laboratory. This report focuses exclusively on Laboratory-owned resources and does not include the resources provided by NERSC, a national user facility. With the fast-changing information technology (IT) and funding conditions, mid-range Computing is an evolving concept. This report presents specific recommendations for Laboratory support and services to enhance the productivity of programmatic clusters. An additional intent of this report is to provide a road map that can be used for long-term strategic planning regarding mid-range computing at Berkeley Lab.

Table of Contents

1. Introduction

In early 2001, members of the Computing and Communications Services Advisory Committee (CSAC) representing the scientific divisions at Berkeley Lab and members of the Information Technologies and Services Division (ITSD) formed a working group that has been actively assessing the need and viability of enhanced Mid-Range Computing (MRC) resources at the Laboratory. The rationale for this working group was the possibility that an easily accessible, high-performance computing facility or resource could become a key component of scientific research at Berkeley Lab.

Mid-range computing at LBNL has undergone significant changes over the last three decades in step with the many changes in the scientific mission of the Laboratory. Since the 1970s, computing at Berkeley Lab has evolved from a centralized facility to decentralized, desktop-centric computing today. This change is a direct reflection of the changing funding situation of the Laboratory from predominantly large groups and block funding to a large number of small groups and individual PIs with smaller grants, as well as the availability of much more powerful and affordable technology.

With the 1995 move of the National Energy Research Scientific Computing (NERSC) center to LBNL, the Laboratory has identified high performance computing as part of its long-term strategic planning. To facilitate the migration of LBNL scientists to high-performance computing, Berkeley Lab invested in a three-year memorandum of understanding (MOU) with NERSC and DOE to provide a fraction of NERSC's computing resources to LBNL users. This MOU has benefited LBNL researchers in a very cost-effective way by providing relatively easy access to one of the most advanced high-performance computers. This very successful three-year agreement, which ended in 2000, represented the only Mid-Range Computing resource planning at the Berkeley Lab since the mid-1990s. Although mid-range computing is identified as key to advancing science, the Laboratory has not had a clearly defined road map or strategic plan in this area.

Among the challenges in such planning is the diverse nature of the research portfolio at LBNL. This scientific diversity imposes a number of constraints on any effort to collectively enhance high-performance computing. With these conflicting parameters in hand, the MRC working group set out a process, described in this report, to assess current computing needs. This process was directed at identifying a common ground and solutions that would most benefit the scientific mission of the Laboratory.

This work, consisting of assessment, findings and determining the appropriate path forward, constitutes the MRC process as described in the following section. A major part of this process includes holding a Lab-wide workshop to assess the needs and independent directions taken by various programs using high-performance computing. In section 3 we will discuss the findings and the new directions that emerged at the MRC Workshop. We will summarize in section 4 where the MRC process stands and describe

the current and future directions of MRC at the Lab, as well as the recommendations for support of programmatic clusters.

2. The MRC Process

Last fall, the MRC Working Group produced a document summarizing the initial assessment and findings of the group on how mid-range computing has been and is being conducted at the Lab, what mid-range computing resources are available at other DOE laboratories, and possible financial models for supporting such resources. This document can be found in the Appendix. Over the last several months, the scope of the MRC Working Group has expanded to investigate Linux clusters. This broadening of the MRC process was triggered by the growth of the number of clusters on site, and the recognition that they provide a cost-effective computing resource. Individuals and programs have followed this Linux path as the answer to their need for mid-range computing. The MRC Working Group started considering what the Laboratory should do to make this approach more successful; in particular looking at what kind of support should be provided and the costs involved.

This second phase of the process culminated with a Mid-Range Computing Workshop that gathered both scientists and computer sciences staff to discuss MRC at the Laboratory. Prior to the workshop, the working group conducted a survey of current and potential users of mid-range computing resources (mostly computer clusters) to assess the status and possible futures of MRC at the lab. The outcome of the workshop and the results of the survey helped define a road map for advancing mid-range computing at the Laboratory (see Section 4.).

3. Workshop and Findings

The goal of the workshop was to evaluate and determine the best path forward for scientific mid-range computing at Berkeley Lab by bringing together current and potential MRC users, for a discussion of MRC users requirements and needs, options and identified offerings. The presentations and minutes of the workshop, as well as the pre-workshop survey results, can be found in [1, 2]. Table 1 is a list of participants, showing the wide representation of all the Lab's scientific divisions and the Computing Sciences directorate. Two areas were clearly identified for discussion at the workshop: a) exploring the feasibility, interest, and scientific impact of a Lab-wide and shared resource, and 2) support for existing and future clusters.

Table 1. List of the MRC Workshop Participants.

Last Name	First Name	Division
	Committee Members	
Adams	Paul	Physical Biosciences
Bashor	Jon	Computing Sciences
Belkacem	Ali	Chemical Sciences
Ciocio	Alessandra	Physics
Downing	Kenneth	Life Sciences
Jung	Gary	Information Technologies & Services
Leighton	James	Information Technologies & Services
Merola	Sandy	Information Technologies & Services
Staples	John	Accelerator & Fusion Research
Tonse	Shaheen	Environmental Energy Technologies
Van Hove	Michel	Materials Sciences
Welcome	Tammy	NERSC
	Attendees	
Shadwick	Bradley	Accelerator & Fusion Research
Ryne	Rob	Accelerator & Fusion Research
Fawley	Bill	Accelerator & Fusion Research
Grote	Dave	Accelerator & Fusion Research
Robin	David	Accelerator & Fusion Research
Lester	William	Chemical Sciences
Head-Gordon	Martin	Chemical Sciences
Miller	Norman	Earth Sciences
Lau	Peter	Earth Sciences
McClung	Ivelina	Earth Sciences
Abadie	Marc	Environmental Energy Tech
Finlayson	Elizabeth	Environmental Energy Tech
Revzan	Ken	Environmental Energy Tech
Dernburg	Abby	Life Sciences
O'Keefe	Mike	Materials Sciences
Zhuang	Vera	Materials Sciences
Ng	Esmond	NERSC
McCurdy	Bill	Computing Sciences
Tull	Craig	NERSC
Chan	Yuen-Dat	Nuclear Science
Hjort	Eric	Nuclear Science
Cromaz	Mario	Nuclear Science
Luk	Kam-Biu	Physics
Spitzer	Chris	Physics
Aldering	Greg	Physics
Spadafora	Tony	Physics
Carithers	Bill	Physics

Support	
J. Hules	CSD
M. Treleaven	ITSD
Y. Mankin	CSD
G. Kurtzer	ITSD

3.1. MRC as a shared resource

One of the approaches initially considered by the MRC Working Group was to create a centralized, shared mid-range computing resource to serve projects in the scientific divisions. However, LBNL is a very diverse research community, and it is very difficult to build consensus on a shared, centralized resource. One option to create such a central shared resource would be to fund the system from Lab overhead funds, as the initial cost is too high to be supported by individual programs. Because of the relatively large investment needed, this option would also require a strategic planning effort at the Laboratory level, involving both scientists and senior management. A strong scientific case would have to be built prior to funding such a resource. In view of the diversity of the scientific programs at the Laboratory, the effort and time required to put together a scientific case for a centralized resource can be large. Since the current status of funding for such a resource is not favorable, it was decided to postpone such a study.

A second option considered by the working group was to encourage projects with existing or planned clusters to pool their resources, thereby creating a more powerful resource than could be procured independently. Although some participants in the workshop saw this as an attractive approach, it became evident from discussions at the workshop that most participants did not see this as a workable idea. Again, the Lab's scientific diversity, and the fact that clusters are often "tuned" to achieve optimal performance for a single application, were identified as stumbling blocks. It was not obvious how to build a resource that would support the range of intended applications needed by different programs. Current and planned owners of clusters will not likely give up the flexibility of individually owned machines if the shared system is not a true MRC that represents a major leap in terms of computer power as compared to their own systems.

As discussed in the working group's report written in Fall 2001 (see Appendix), an MRC system that goes beyond a pooling of resources cannot succeed without a strong scientific case and a commitment of support from individual programs, divisions and the Laboratory as a whole.

3.2. Support of individual clusters

The Lab has seen rapid growth of the number of computer clusters. From the hardware perspective clusters represent an affordable solution. However, setting up and running an efficient cluster system is neither trivial nor cheap.

There is a general consensus that providing expertise for people buying clusters would be useful. Ideally, the Laboratory should have experts (from ITSD) who could provide pre-purchase consulting on hardware and software. These experts would be able to advise which machines to buy, and possibly leverage volume buying into better pricing. It is clear that the more scientists buy different systems, the more difficult it is to provide efficient and cost-effective central support to these systems.

Clusters are set up at various locations with mixed infrastructure/environment qualities. Providing space for clusters would be a valuable service. This will create a machine room environment with easy access to electrical infrastructure, proper air conditioning and access to high-speed local area and wide area networks.

In many scientific groups, responsibility for system administration falls on students and postdocs - and doesn't fall evenly. Furthermore, that knowledge disappears when the student or postdoc moves on to other horizons. A centralized system administration would constitute a very attractive solution, ideal for ensuring peak performance and availability. It would also help minimize the cybersecurity vulnerabilities of a system. However, such support is often perceived as prohibitively expensive. ITSD should explore how system management support can be provided at more affordable cost levels, perhaps by tailoring support to specific needs and offering a range of support levels.

Parallel programming is perceived as a major barrier for several researchers who contemplate following the path of clusters. However, there is a lot of programming knowledge at the Laboratory, particularly in the NERSC Center Division, that can be brought together to help current or future cluster users. A path forward here is to encourage scientist-to-scientist help and support. This can possibly be achieved by creating a web site and users group where knowledge can be shared.

4. Summary and Path Forward

Although the workshop demonstrated that the need for mid-range computing is clearly identifiable, due to disparities of the individual needs as well as a perception that no substantial economies of scale would be realized by pooling resources, a mid-range computing resource as a more substantial institutional resource remained an open issue. This step up in the availability of an LBNL-owned major computer resource needs both a strong scientific case and a commitment from the Laboratory management to become viable. There are also difficulties associated with the high initial costs of acquiring and operating such a resource.

The workshop also identified need for pre-purchasing support and possibly computer room space, and the desire for affordable system management. As a result of the workshop, a well-defined path forward was identified in the area of support of individual clusters.

The current effort of the MRC process is on a proposal for a three-year start-up Scientific Computing Support Program that would provide various services to selected projects. Projects that applied to be included in the program will go through a competitive review process based on the science, the budget, the time frame and a broad representation of most of the scientific divisions. The proposed program would include existing and planned clusters and is intended to reduce costs to scientific programs of Berkeley Lab.

From the user's perspective it would not require secondary staffing coverage for a non-standard, non-optimal, and/or difficult to maintain configuration and the scientific staff turnover would not be an issue. By providing professional system administration support, scientific staff will be able to focus on their work on science.

While a strong scientific justification would be needed to secure overhead funding for this program, the three-year Support for Scientific Computing program would undoubtedly add value to the scientific programs of Berkeley Lab.

Independently of the outcome of this proposal, this program should not preclude an institutional mid-range computing system at some future point if more opportune times are present both from the perspective of strategic needs of the Laboratory as well as from the perspective of funding.

References

[1] http://www-atlas.lbl.gov/~ciocio/CSAC/MRC/Workshop/presentations/
[2] http://www-atlas.lbl.gov/~ciocio/CSAC/MRC/Workshop/proceedings/Minutes.doc

Appendix. The MRC White Paper

An Institutional Scientific Mid-Range Computing Resource for Berkeley Lab

A report compiled by the Mid-Range Computing Working Group of the Computing and Communications Services Advisory Committee and the Information Technologies and Services Division:

Paul D. Adams, Physical Biosciences
Jon Bashor, Computing Sciences
Ali Belkacem, Chemical Sciences
Alessandra Ciocio, Physics
Kenneth H. Downing, Life Sciences
Gary Jung, Information Technologies and Services
James F. Leighton, Information Technologies and Services
Alexander "Sandy" Merola, Information Technologies and Services
Douglas L. Olson, Nuclear Science
John W. Staples, Accelerator and Fusion Research
Shaheen Tonse, Environmental Energy Technologies
Michel A. Van Hove, Materials Sciences
Tammy S. Welcome, NERSC

Executive Summary

As the role and contributions of high-performance computing continue to increase in significance, Berkeley Lab scientists are seeking out potential advantages provided by more powerful computing resources. These resources range from small clusters developed independently by Lab groups to such high-performance systems as those provided by NERSC.

Based on these indicators, a CSAC-ITSD working group has investigated whether an institutional mid-range computing resource would be appropriate and/or sustainable for Berkeley Lab. This report represents the culmination of the first stage of the group's work. The working group has identified various options for implementing an institutional mid-range computing resource and identified related financial considerations. The next step is to initiate discussions of such a resource with senior Lab management and the pool of potential users at the Laboratory. Those discussions, together with the information already collected, will then determine the appropriate path forward.

Is an Institutional Mid-Range Computing Resource Appropriate for Berkeley Lab?

The Laboratory's Computing and Communications Services Advisory Committee (CSAC) and Information Technologies and Services Division (ITSD) are working in partnership to determine whether there is sufficient institutional value in procuring a Lab-wide, mid-range computing resource as a tool for scientific research.

Scientists today have access to desktop workstations more powerful than high-performance computers (HPC) of 25 years ago; many research areas can benefit from today's increased HPC power. At Berkeley Lab, HPC usage has typically meant scaling up to increasingly powerful computing systems, including the use of NERSC resources. However, there is still a wide gap in terms of computing power and architectures between desktop workstations that are generally available to Berkeley Lab researchers and large-scale, high-performance computers. One option for bridging this gap is to have a resource that is "mid-range" between workstations and high-performance computers similar to those operated by NERSC. Cluster computers may offer a potentially attractive and cost-effective mid-range computing option.

Berkeley Lab currently lacks such a generally available computing capability, and some Berkeley Lab researchers have shown an increasing interest in the potential of such a resource. This interest can be seen in the growing number of small clusters of computers assembled by groups in various scientific programs, the purchase of larger off-the-shelf clusters by several groups, and the growing number of Berkeley Lab scientists who are applying for and being allocated computing and storage resources from NERSC. A good example of a mid-range computing resource at the Lab is the Parallel Distributed Systems Facility (PDSF), a 281-processor cluster currently being significantly upgraded. PDSF is used primarily by researchers in the Nuclear Science and Physics divisions.

A working group made up of CSAC and ITSD members has been assessing whether there is sufficient need and support for such an institutional resource among Berkeley Lab researchers, and to identify additional investments, if any, that Berkeley Lab should make in mid-range computing capabilities. Among the options discussed to date are:

1) Providing access to the Lab's newly installed 160-processor cluster named "alvarez," perhaps with an upgrade
2) Contracting for access to computing resources from NERSC, as was done under a special three-year program
3) Procuring an additional computing resource
4) Outsourcing mid-range computing resources
5) Making no change at this time

These options will be described in more detail in a subsequent section.

How the Working Group Is Proceeding

The CSAC-ITSD working group has been investigating the potential of an institutional mid-range computing resource for Berkeley Lab since early 2001. Mid-range computing at Berkeley Lab has a mixed track record – there have been both high-profile failures and low-profile successes – so the group is committed to making a thorough investigation before coming to any conclusions or making any recommendations. As part of the group's investigation, which began in early 2001, we have gathered data on:

- How mid-range computing has been and is being done at LBNL;
- What, if any, mid-range computing resources are available to scientists at other DOE laboratories (this data is included in the appendix); and
- Possible financial models for supporting such a resource.

The next phase of the group's work is to identify potential users (taking into account both scientific suitability and ability to contribute funding) of a mid-range computing resource within the Lab's research community. The group will then hold focused workshops with these potential users to research the various needs and refine the systems specifications and accompanying financial model.

Two Critical Components for Success

The rationale behind the discussions and ongoing effort on MRC is the possibility that a generally accessible high performance computing facility could become a key component of scientific research at Berkeley Lab. To date, the Lab has not clearly defined a plan to broadly integrate scientific computing into Lab programs, although significant investments have been made in this direction. In charting a future course, two separate but essential issues must be addressed.

The first issue is usefulness. To be useful and succeed, the mid-range computing facility:

a) Should respond to the needs of a broad range of users.
b) Should provide a computing resource that is significantly more powerful than a system that an individual researcher or group could obtain. It should be readily available, it should have a high turnaround rate, it should have a configuration that responds to the needs of users and it should be relatively easy to use.
c) Should be perceived by a scientist owning a small cluster as a major step up in terms of advanced computing power and software.
d) Should be upgradeable—and upgraded regularly to keep up with advances in technology
e) Should be much more cost-effective than owning a small cluster.
f) Should be operated in an expert manner.
g) Should be responsive to user needs, requests and input.

The second issue is commitment. There should be a clearly expressed need by scientists (and concomitant involvement), a strong commitment from the scientific divisions, and a strong commitment from Lab management.

These requirements will put a major burden on the design, operation and sustained funding of a mid-range computing facility even when a strong need is identified. It will be very difficult to immediately achieve this goal and a gradual approach may be more appropriate. Since Berkeley Lab will be starting from the ground floor in terms of running a generally available HPC facility, it could be a few years before such a resource is running seamlessly.

History and Current Status of High-Performance Computing at Berkeley Lab

High-performance computing has been a component of Berkeley Lab research since the 1960s. The first supercomputer ever connected to ARPANET was a Control Data Corp.

6600 located at Berkeley Lab. In the mid-1970s, when the Magnetic Fusion Energy Computer Center (NERSC's predecessor) was just being launched at LLNL and consisted of one oversubscribed CDC 7600, jobs beyond the computer's capacity were driven to Berkeley Lab to be run overnight, with the results couriered back to Livermore in the morning.

In 1993, before NERSC arrived, Berkeley Lab installed a high-performance computing resource, the 4,096-processor MasPar MP-2 supercomputer. Unfortunately, many Lab scientists found it too difficult to make the transition to parallel programming and an unfamiliar operating system, and the system was out of service by the time NERSC arrived in 1996. Lessons learned from this experience include the need to provide strong user support and the need for a viable financial model to provide ongoing funding.

Among the benefits expected to accrue from NERSC's move to Berkeley Lab was an increase in the role of computational science among Lab research efforts. On an institutional scale, this goal has been achieved, as indicated by a remarkable number of scientific achievements using HPC as an underlying technology, the growing number of Berkeley Lab researchers being allocated time on NERSC's supercomputers, a separate three-year program to provide a portion of NERSC's Cray T3E for the exclusive use of Lab and UC Berkeley scientists, and the establishment of a Computational LDRD program at the Lab.

LBNL Users of NERSC

As employees of a DOE national laboratory, Berkeley Lab scientists have been able to apply for time on NERSC systems since the 1980s, and since NERSC moved to LBNL in 1996, Berkeley Lab users have accounted for about 10 percent of the total usage of the parallel systems. Through several Lab-based efforts described in this section, Berkeley Lab has become one of the top institutional users of NERSC, moving from being ranked eighth on the list of institutional allocations to being ranked third in the five years since the center was relocated.

Berkeley Lab Investments in HPC

Berkeley Lab has used University of California funds to invest in two hardware systems. In 1997, Berkeley Lab secured a 3.2 percent augmentation of the Cray T3E and committed to three years of support. This investment leveraged the original NERSC-2 system and provided separate allocations for Berkeley Lab users. This provided several scientists who were new to parallel computing an opportunity to learn how to exploit NERSC computational resources. Although the number of users varied from year to year, the number of hours allocated through this effort nearly doubled from year to year, as shown below.

Fiscal Year	Number of Allocations	Total Hours Allocated
FY98	12	50,000
FY99	18	95,000
FY00	13	191,500

In FY2000, Berkeley Lab added to this investment by again using University of California funds to acquire a 160-processor PC cluster (named "alvarez"), currently

managed by NERSC. The Lab has made a commitment for ongoing financial support. Initially, this cluster will provide a dedicated HPC resource for a few strategic projects, as well as serve as a platform for computer science R&D conducted by NERSC. In subsequent years, the cluster may become available for a wider range of users, but plans for this transition are not yet in place.

Computational LDRD Projects

To foster and improve computational science projects across all Berkeley Lab divisions, the Lab created a computational science Laboratory Directed Research and Development (LDRD) program in FY1996. The goals in creating this program were to bolster LBNL's use of high-performance computing in all disciplines and to make scientific computing a "core competency" of the Laboratory. This effort represented a significant investment by the Lab—about $3 million over the first three years.

In the first phase of the program, from 1996 to 1999, about 20 projects were funded, which brought about 20 postdocs to the Lab and trained many students in computational science. In the second phase, from 1999 to 2001, the program focused on large-scale teams and strategic collaborations.

The program successfully advanced the role of computational science as a component of research at Berkeley Lab (a factor that contributed to the motivation for this investigation). Over the five-year period it helped to increase significantly the share of NERSC allocations going to researchers at Berkeley Lab. Whereas in FY1996 Berkeley Lab ranked only eighth on the list of institutions receiving allocations at NERSC, in FY2001 Berkeley Lab moved up to third place (after LLNL and ORNL).

Interestingly, some of the most significant scientific achievements of Berkeley Lab scientists in the past few years were Berkeley Lab projects seeded by this LDRD program and which used NERSC as a computing resource. The Supernova Cosmology Project (Perlmutter), the complete solution to breakup of a quantum system of three charged particles (McCurdy), and the analysis of the BOOMERANG experimental data (Borrill) to determine the geometry of the universe not only were NERSC-utilizing projects involving "graduates" of the computational LDRD program at Berkeley Lab, but their results also made the covers of *Science* and *Nature* magazines.

These programs demonstrate that Lab researchers are benefiting from access to large-scale computing resources and that such resources are significantly contributing to the quality of science at the Lab.

Other Bigger-Than-a-Desktop Computing Efforts

An informal survey of the Lab conducted as part of this investigation has found a handful of cluster computer systems being used by individual research programs. Clusters are assemblies of commodity computers designed and networked to operate as a single system. By using off-the-shelf components, clusters may provide a cost-effective balance between price and computer performance. These systems can either be assembled from individual computers or purchased as "plug-and-play" assemblies complete with software. Clusters are used by the following groups:

- The Supernova Cosmology Project, which uses a 5-node cluster for about 10 users;

- The Yucca Mountain Project, which has assembled a 10-node cluster and plans to add six more nodes;
- The Center for Computational Geophysics, which has purchased an 8-node Linux cluster;
- The Berkeley Drosophila Genome Project, which has purchased a 20-node Linux cluster and is adding 12 more nodes;
- NERSC's Future Technologies Group, which has operated 12-node and a 32-node research clusters and develops software to improve the performance of Linux-based clusters.

A graphical representation of the various cluster and high-performance computing systems at Berkeley Lab is included as an appendix to this report.

PDSF—A Mid-Range Computing Success Story

In 1996, a collection of HP, Sun and SGI workstations orphaned by the cancellation of the Superconducting Supercollider arrived at Berkeley Lab. The system, known originally as the Particle Detector Simulation Facility, was rechristened the Parallel Distributed Systems Facility (PDSF) and dedicated to supporting high energy and nuclear physics research. Since then, the hardware and software has been constantly upgraded from a few dozen processors, and today the PDSF is a Linux cluster consisting of 281 processors with a theoretical peak processing capacity of 155 gigaflop/s and a total storage capacity of 7.5 terabytes. The system has also been upgraded with additions of high-bandwidth networking, disk cache and interoperability with NERSC's High Performance Storage System (HPSS). PDSF is used primarily for the STAR experiment, but also currently supports 13 other projects.

PDSF is run as a partnership between Physics, Nuclear Science and NERSC Divisions, with all three divisions contributing to cover the costs of four NERSC employees who provide system administration and user support. The user community funds expansion and upgrades of the system. When users want to expand the system, a portion of the cost of the upgrade is used to make accompanying improvements in the overall computing and networking infrastructure.

This cooperative model has worked well for PDSF, allowing the system to provide a reliable, well-supported resource and to expand to meet users' changing needs.

What Are Berkeley Lab's MRC Options

If the decision is made to pursue the addition of a mid-range resource, the partnership example set by PDSF may provide a viable model. As shown by the PDSF project, such a resource can be obtained, operated and upgraded by Lab divisions, with the institution providing the needed infrastructure, perhaps a startup subsidy, and ongoing support.

Another lesson from PDSF is that an existing resource can be adapted for use by a larger number of projects. As mentioned earlier in this document, the working group has identified four possible paths forward, as well as the option of making no change at this time. Here is a discussion of those options.

Providing access to the Lab's newly installed 160-processor cluster, "alvarez"

Berkeley Lab has purchased and installed a 160-processor IBM cluster computer (named "alvarez" after Lab Nobel Laureate Luis Alvarez) so that NERSC can assess whether such a system can meet the heavy day-to-day computing demands of a broad range of scientific projects. To date, most large scientific commodity clusters are used more for specific research applications than as resources shared by a number of projects in a variety of scientific disciplines. Such a resource must also be robust enough to be consistently available to meet user demand.

Another objective of "alvarez" is to provide a computational resource for strategic Berkeley Lab projects and campus collaborations that require significant computational resources. The experience gained in using a cluster to support Lab research could lead to more extensive, cost-effective computational offerings in the future. After NERSC completes its evaluation of the cluster, it may be available to a wider range of Lab users.

Contracting for access to computing resources from NERSC

This approach taken in FY98-00 was described earlier in this report. This model could be used again, as long as NERSC is able to provide the resources, and institutional funding

can be procured. It would also be useful to conduct an evaluation of the previous program's successes and limitations, if this model is chosen.

Procuring an additional computing resource

This option would allow the Berkeley Lab to design and implement resources specifically tailored to meet the needs of Lab researchers, as opposed to trying to adapt existing hardware. This approach requires a large initial level of investment to ensure that adequate hardware and software resources are obtained at the outset.

Outsourcing additional computing resources

Although Berkeley Lab has traditionally operated its own computing systems, outsourcing mid-range computing may be an option. (The most recent example of the Lab outsourcing institutional computing was the use of a vendor to support legacy codes that could only be run on an obsolete IBM mainframe, and this was the most cost-effective approach.) This would be a new model and would require substantial study before proceeding.

Making no change at this time

Clearly, if there is insufficient scientific interest and/or inadequate funding an institutional computing at this time, the idea could be postponed and perhaps revisited should circumstances change.

A Financial Model for Institutional Mid-Range Computing

Should the need and/or demand for an institutional mid-range computing resource be identified, the next – and perhaps most challenging – steps will be to find both a technical solution and a financial model that will work. Providing this computing resource will require a substantial investment by the Laboratory and this investment has to be sustained over the lifetime of the system. The financial model developed to support this resource must also include provisions to protect the funding from fluctuations in DOE budgets. The financial model must take into account the fiscal realities of Berkeley Lab.

- The Lab has been reducing overhead and this trend is not likely to be reversed. Using overhead funds to pay for a project is tantamount to recharging all Lab programs, so there is likely to be strong resistance to an ongoing use of overhead funding to pay for an institutional mid-range computing resource, especially as this would be a multi-year commitment.
- On the other hand, relying to a large degree on recharge to fund the operation and upgrades of a facility (after it has been purchased) does not appear to be viable, as this mechanism was one of the factors contributing to the demise of the MasPar computer.
- Some scientific divisions within the Lab already spend a substantial portion of their budget on scientific computing (hardware, software and support) every year. Hardware is usually purchased as capital equipment, decided upon at the division director level. The other costs are often hidden in that they are made piecemeal or covered by the salary of employees who do support as a sideline. To provide an attractive alternative, a mid-range computing resource would have to be significantly more powerful than a system that could be procured at the division level and the associated support costs would have to be shown to be reasonable.

A Viable Financial Model

We propose that a viable financial model would involve strong commitment (and funding) up front from at least several scientific programs and divisions, in conjunction with a contribution from Lab overhead funds. A plausible scenario would be to create a facility that essentially belongs to the scientific divisions and is configured with input from the users. Operation and system management would be funded through overhead and would be provided by the computing support component of ITSD. Having the system centrally managed would benefit the supporting divisions by relieving them of responsibility for operation and management, software, maintenance costs and cybersecurity. The option of leveraging NERSC resources could also be explored.

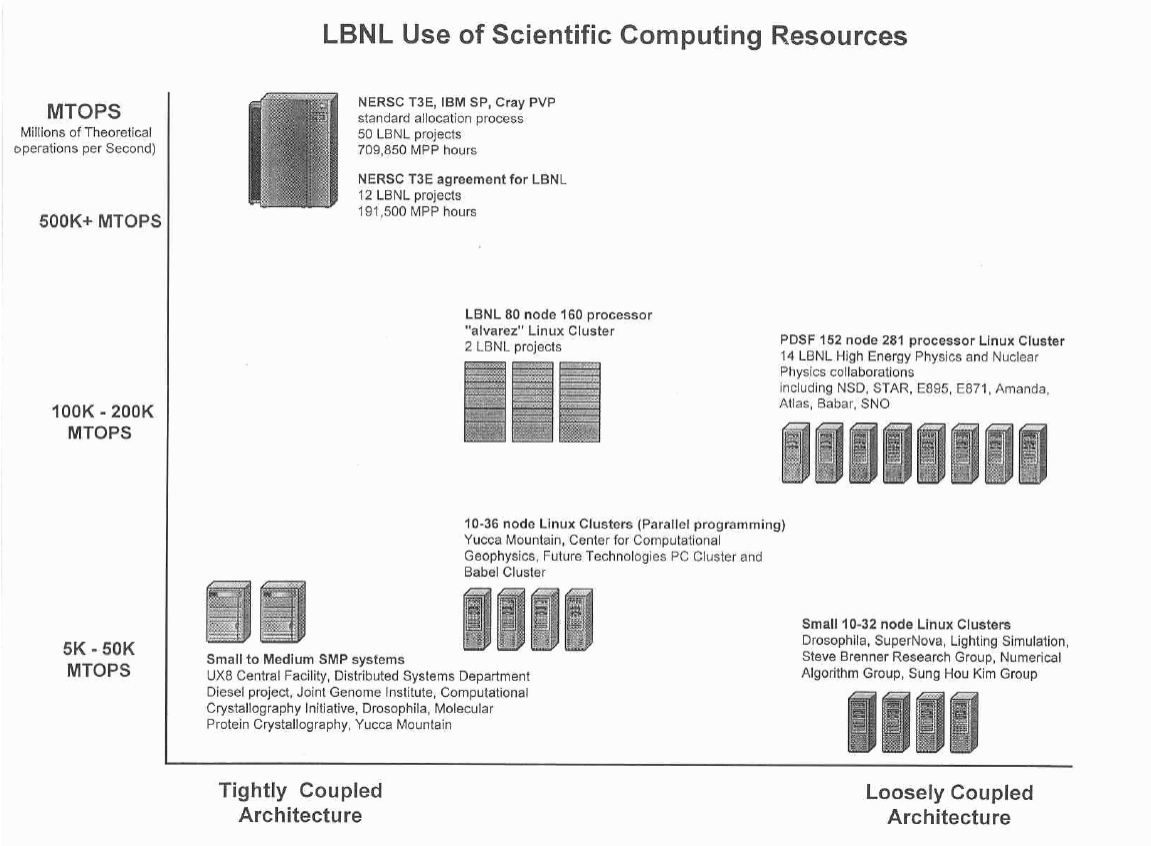
Divisions supporting the system with funding would receive use of the resource in proportion to their financial support. Divisions that don't buy in could still have access to the resource, but on a recharge basis.

Detailed budget estimates for several options considered by the working group are included in the appendix section of this report.

Supplemental Materials

1. LBNL Use of Scientific Computing Resources
2. Mid-Range Computing Budget Estimates
3. A Survey of Mid-Range Computing Resources at Other Labs

1. LBNL Use of Scientific Computing Resources



	Division	Project	Nodes	Model	CPU
Small 10-32 node Linux Clusters	Genome	Drosophila	20	Intel	2x 700Mhz PIII
	Physics	SuperNova	5	Intel	6 nodes @ 300Mhz; 3 nodes @ 600Mhz PII
	Energy and Environment	Lighting Simulation	10	AMD	1.2Ghz Athlon
	Physical Biosciences	Steve Brenner	12	Intel	2x1Ghz PIII
	NERSC	Numerical Algorithm	12	AMD	1.3Ghz Athlon
	Physical Biosciences	Sung Hou Kim	6	Intel	2x1Ghz PIII
10-36 node Linux Clusters (Parallel programming)	Earth Sciences	Yucca Mountain	20	Intel	2x1Ghz PIII
	Earth Sciences	Center for Computational Geophysics	8	Intel	2x1Ghz PIII
	NERSC	Future Technologies PC Cluster	36	Intel	400Mhz PII
	NERSC	Future Technologies Babel Cluster	12	Alpha	466Mhz Alpha
Small to Medium SMP systems	ITSD	UX8 Central Facilities	1	Sun E4500	8 ea. 400Mhz Ultrasparc II CPUs
	Distributed Systems	Diesel project	2	Sun E4000	8 ea. 400Mhz Ultrasparc II CPUs
	Distributed Systems	Diesel project	2	Sun E4500	4 ea. 400Mhz Ultrasparc II CPUs
	Genome	Joint Genome Institute	1	Sun E6500	20 ea. 360Mhz Ultrasparc II CPUs
	Genome	Joint Genome Institute	1	Sun E3000	8 ea. 400Mhz Ultrasparc II CPUs
	Genome	Joint Genome Institute	1	Sun E450	4 ea. 400Mhz Ultrasparc II CPUs
	Physical Biosciences	Computational Crystallography Initiative	2	Compaq ES40	4x833Mhz Alpha
	Physical Biosciences	Computational Crystallography Initiative	1	Compaq DS10	466Mhz Alpha
	Physical Biosciences	Computational Crystallography Initiative	1	Compaq XP900	466Mhz Alpha
	Physical Biosciences	Computational Crystallography Initiative	1	Compaq DS20E	2x667Mhz Alpha

2. Mid-Range Computing Budget Estimate

	Option 1 - alvarez			Option 1a - alvarez+			Option 2 - New Cluster			Option 3 - SMP		
	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3	Year1	Year2	Year3
MRC Purchase (no overhead)	0			500			700			1200		
plus procurement burden 3.9%	0			19.5			27.3			46.8		
plus materials handling 4.2%	0			21			29.4			50.4		
Procurement Total	**0**			**540.5**			**756.7**			**1297.2**		
Vendor support												
HW Maintenance (8x5) and SW Maintenance	0	105	105	0	240	240	105	105	105	180	180	180
Vendor Support Hotline	25	25	25	25	25	25	25	25	25	25	25	25
Additional software												
3rd party tools and applications												
Permanent license	281			338			281			129		
Annual software maintenance	50	50	50	65	65	65	50	50	50	30	30	30
Subtotal Vendor support and software	356	180	180	428	330	330	461	180	180	364	235	235
plus procurement burden 3.9%	13.88	7.02	7.02	16.69	12.87	12.87	17.98	7.02	7.02	14.20	9.17	9.17
Hardware and Software Support Total	**370**	**187**	**187**	**445**	**343**	**343**	**479**	**187**	**187**	**378**	**244**	**244**
Procurement Team Effort												
6 staff 0.5 FTE for 6 months (procurement, technical, benchmarks, etc…)				75			225			225		
Facilities Costs												
Base installation costs (seismic design, bracing, wiring)				20			60			60		
Power Distribution Unit including installation							60			60		
UPS including installation							60			60		
Space/Electricity	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6
Staff Support (incl payroll and org burden)												
System Administration 1 FTE, Project Management 0.5 FTE	225	236	248	225	236	248	225	236	248	225	236	248
User Services 0.5 FTE	75	79	83	75	79	83	75	79	83	75	79	83
Applications Assistance 1 FTE	150	157.5	165.4	150	157.5	165.4	150	157.5	165.4	150	157.5	165.4
Total incl purchase, vendor support, software, staff and facilities	**823**	**663**	**687**	**1459**	**819**	**843**	**2094**	**663**	**687**	**2534**	**720**	**744**
Subtotal	2174			3121			3444.4			3998.4		
25% planning margin	543			780.2			861.1			999.61		
Estimated 3 yr. Total Cost of Ownership	**2717**			**3901**			**4305.5**			**4998**		

Option 1 - alvarez

NERSC may release the LBNL alvarez cluster in late 2002 or early 2003. By this time, the alvarez hardware will be two years old, but still very usable. Costs for this option will include hardware maintenance, new software purchase to meet the needs of LBNL scientific programs, software maintenance, facilities costs and staff for systems administration and consulting.

Option 1a - alvarez+

Same as option 1 except that LBNL will purchase additional nodes to expand cluster. Software purchase and maintenance costs are higher because of more nodes.

Option 2 - Purchase new cluster

LBNL decides to acquire new cluster system to meet MRC needs. Costs include purchase cost + procurement burdens, procurement team effort and support costs as described in option 1. Purchase costs are lower than for SMP system outlined in option 3, but software licensing and staffing costs are higher.

Option 3 - Purchase new SMP system

LBNL decides to acquire new SMP system to meet MRC needs. Costs include purchase cost + procurement burdens, procurement team effort and support costs as described in option 1. Purchase cost is higher for an SMP system than a cluster system; however software costs are about half as much because licensing is typically per cpu. Systems administration and consulting costs are lower.

3. A Survey of Mid-Range Computing Resources at Other Labs

As part of this study, the CSAC/ITSD working group surveyed other national laboratories to determine whether institutional MRC resources were available and, if so, how the resource was managed and supported. This informal survey found that mid-range computing is a mixed bag at other labs, but did provide useful information should Berkeley Lab seek to provide such a resource.

Of the 11 national laboratories investigated, only Argonne National Laboratory, Oak Ridge National Laboratory, and Lawrence Livermore National Laboratory (LLNL) have some semblance of institutional (lab-wide, lab-accessible) MRC. Of these three, only LLNL has MRC as a lab-wide resource. Although used for unclassified computing, LLNL MRC is operated in conjunction with the Stockpile Stewardship Program and benefits from investments made in the unclassified Accelerated Strategic Computing Initiative (ASCI), as well as other previously existing infrastructure.

Most (85 percent) funding for LLNL MRC comes from LLNL (i.e., institutional funding). Mechanisms exist for user programs to contribute funds to the central computing facility. This program-provided funding amounts to only about 15 percent of the total funding, but is considered important for community buy-in and accountability. Programs provide additional funds either as block funding (a contract for a certain amount of resources) or as co-investment (with funding added to equipment procurements and programs receiving appropriate resource allocation). Berkeley Lab's PDSF operates similarly to the co-investment model.